Integrated Information Networking Infrastructure

Background and Rationale

As emphasized in the Administration's "Agenda for Action" white paper, perhaps the most significant technological shift in this decade will be in the way people deal with information, from the personal level, business level, and the government level. Information has become one of the nation's most critical economic resources, for service industries as well as manufacturing, for economic as well as national security. In an era of global markets and global competition, the technologies to create, manipulate, manage and use information are of strategic importance for the United States. Those technologies will help U.S. businesses remain competitive and create challenging, high- paying jobs. They also will fuel economic growth which, in turn, will generate a steadily-increasing standard of living for all Americans. For all these same reasons, these technologies will also be of strategic importance in maintaining the Laboratory's ability to be both stewards and providers of advanced weapon technologies.

Technologically, the NII consists of basically of three general components: hardware infrastructure (primarily the network), software infrastructure (primarily software services), and the applications (working interoperably and over the network). The NII will consists of the integrated working together of these components for everyone's benefit. "*Every component of the information infrastructure must be developed and integrated if America is to capture the promise of the Information Age*"

The administration is committed to working with private industry to establish the basic principles of the NII, and to use regulation to encourage private sector investment, technological innovation, to ensure universal service, to encourage interoperability, and to ensure information security and trustworthiness.

The strategy we are choosing to meet this technological challenge is to develop an integrated approach to applying the information technology to science and society. This includes using requirements from a suite of applications to define a common infrastructure. We then will develop an array of technologies to meet these requirements. This process allows, for example, a solution of problem of importance to health-care to rapidly impact problems in industrial manufacturing, or even gas and oil exploration and production. This involves not only finding common solutions but in tightly coupling the research and development and information management so that these new tools can be more rapidly deployed both in the laboratory and by industry, itself. More information can be obtained through the World Wide Web at the URL: "`http://www.acl.lanl.gov/sunrise/sunrise.html`".

Program development opportunity

The four components of the Federal NII program include

- Information Access to improve public access to government information by allowing agencies to make such information available electronically ,
- National Performance Review to use information technology investments to move towards *Electronic Government*,
- High Performance Computing and Communications Program,
- and Service Delivery to use information technology to deliver services such as loans, benefits, and grants directly to the public.

The Federal High Performance Computing and Communications program has allowed DOE to continue its leadership in this core area, long supported by the weapons program. This has resulted in a flag ship effort in the national high performance computing arena at Los Alamos. Because DOE is getting no additional funds in FY94 or FY95 for NII activities, the DOE's leadership position is eroding rapidly. In FY95 the President's budget contains over $3 billion for NII activity which we are aggressively pursuing. This includes requesting the inclusion of $50M into the federal budget for the DOE in FY95. The linkage of these various components is crucial to the economic competitiveness of the nation and to the actual use of high-performance computing in its various forms throughout industry. As is clear from the emphasis of the national program, the ability to deal with large amounts of distributed information intelligently and efficiently will be fundamental to not only the success of High Performance Computing but to the societal infrastructure as a whole later in this decade. The potential revenue to the laboratory in this area alone is well in excess of $10 million/year and can leverage at least 10 times that in related efforts in programs that use the infrastructure tools.

By developing an integrated approach to the Information Infrastructure and linking High Performance Computing with commodity software and hardware tools, we have an opportunity not only for Federal funding but also working directly with industry to assist them in their large data management and a analysis problems. This is possible because of our unique position in understanding the computational algorithms and integration technology.

Connection to strategic directions and laboratory goals

The Laboratory has a goal to sustain its world-class strength in High Performance Computing and to work closely with industry on a variety of areas including materials manufacturing and a variety of others. We have identified several industrial partners which are working with us to advance the state of the art in their field. These include manufacturing companies like Gatan that market Tunneling Electron Microscope equipment, to Xerox who is seeking new ways of developing Xerographic engines, to Adobe and HaL who are developing advanced tools for electronic publishing and network navigation, to the National Jewish Center for Immunology and Respiratory Medicine, to the Dept of Commerce.

R&D approach and likelihood of project success

The HIIII project was started in FY94 to enable the laboratory to take an aggressive stance in NII and provide a strong R&D base for obtaining NII funds from various government agencies and even from U.S. industry.

In order to achieve this ,we have set the following three objectives:

- To develop common information-enabling tools for advanced scientific research and its application to industry
- To enhance the capabilities of important research programs at the Laboratory
- To define a new way of collaboration between computer science and scientific research and development.

The basic paradigm being developed involves a document-centric user interface which will allow arbitrary object support including embedded applications, multimedia video/voice fragments and links to a wide information space.

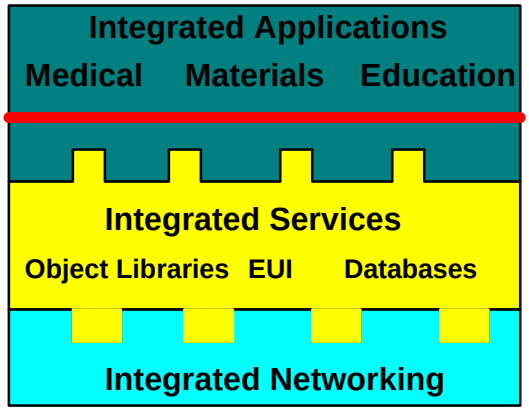
We are developing an "information kiosk or portal" based on an ATM network so that all the participants can exchange, "publish", or interact on applications and data. This will function on heterogeneous platforms, provide for constrained access to data through security mechanisms, and be extensible. The data-mining technology will include the ability to quickly browse large complex image databases with various feature extraction capabilities, provide advanced, selective, compression algorithms, and the ability to merge and purge large complex data sets.

The focus is on real scale problems, reflecting our belief that only experience with real problems encompassing huge data sets will facilitate true progress. Dealing exclusively with toy problems is as likely to mislead as it is to inform. For example, trivial issues like name fusion, can become critical with huge data sets. Therefore, our focus will be on developing prototype tools that can be tested in real-sized testbeds.

We believe the benefits will be to the laboratory as a whole in helping define the way research will be done in the future, in enhancing the competitiveness of laboratory research programs, and in enhancing the nation's ability to use advanced information technologies for applications of importance to industry.

We are taking an integrated approach in which several applications in disparate areas are developed simultaneously. This is to avoid vertical solutions applicable to only one domain, but rather to raise the level of commonality and interoperability to a new level. The identified application areas are at the core of the NII and include telemedicine with the potential for substantial change in the way health care is carried out, materials modeling and analysis which can change the way materials data is analyzed and dispersed in the manufacturing community, transportation [analysis] which has the potential to change the way traffic information is utilized and understood to improve the functioning of the cities, and finally education in which what is being taught as well as the methods of teaching are currently under rapid change. These four areas represent a wide diversity in networking requirements, but all have some need for multimedia data delivery and thus need as wide a bandwidth and efficient protocols as possible.

The role of networking in supporting integrated applications and the integrated services is show below.



We envision the use of the Sunrise environment by novice and expert users who wish to explore the NII and/or collaborate in research. The environment will provide the users with an interrelated set of tools which will enhance their interaction with a domain of knowledge. The interaction of the tools and the knowledge domains might produce a medical diagnostic system, an environment for materials development and design, or a classroom learning situation. The tools will facilitate the formulation of a problem, it's solution, the interpretation of results and the publishing of those results. In support of these tasks, the tools will provide mechanisms to query the range of information sources on the NII, to consult with experts in a knowledge domain, and to collaborate with co-workers in formulating and solving problems. The envisioned toolset has a number of components:

- A journalling facility, similar to a scientist's log book, to track the individual's research progress including the data sets accessed, the results of data analysis, the individual's interpretation of those results, and monitoring the individual's interactions with others. [Note that there are privacy implications here]. The log would be secure, and would provide authenticated time stamps which might be used to validate rights of first discovery of some idea. It will utilize object encapsulation to deal with multiple problems, patients, categories, etc. Appointments, meetings, etc. would be accommodated.

- An interface to experimental equipment in order to initiate, monitor and analyze an experiment connected over the network to the system. Similar functionality for initiating detailed computational simulations on remote supercomputers will also be provided.
- A remote data access and display capability which interacts with various distributed databases included OODB's over the network. It will provide a complex querying interface supporting text, multi-spectral images, and digital video attributes (perhaps combining with the results of previous queries in a hierarchical manner). It provides for the publishing of data by the user and the ability to "purchase" remote data either by value or by reference.
- A digital document repository which facilitates the retrieval and publishing of research documents. This might be just a variation on the more general data access tool. These documents might be hypermedia tours of selected parts of the individual's logbook.
- An accountant which will monitor the cost or potential cost of actions and invoke an authorization process under certain conditions and provide rapid feedback on the actual cost of the actions. (This is a variant of a work-flow capability) Such an authorization might be triggered when a remote data request accesses commercially billed information, which might only be advertised by attributes. The actual data in that case might be the part that costs.
- A toolbox of analysis modules which provides processing of data to facilitate the actual research.
- A facility that can provide a range of help from concise command summaries all the way to live telecollaboration with paid consultants
- User annotations across the entire domain would be supported. These also could be made available to collaborators or published more widely. We envision native viewers of much of the data with more sophisticated domain-specific analysis tools being invoked as needed.

Overview

Stephen Tenbrink and John Morrison, C-5 : $500K

The Sunrise Project will involve the use of several technologies that will enhance the effectiveness of computers in the office environment of a laboratory staff member. A critical part of this effectiveness is communication and information retrieval. Using email and other Internet applications of today as models, we hope to expand the capabilities of a workstation and its network to provide such things as video on demand, teleconferencing, visualization postprocessing, and other tools that free up time for the user to be more effective in other endeavors.

A network technology that promises to support this vision is one were various data rates and qualities of service are supported. The emerging Asynchronous Transfer Mode, or ATM, appears to provide that support and was chosen as the networking component of Sunrise. The plan is to procure three ATM switches and place each switch in a strategic location for researchers to access the Sunrise Network. One obvious site is the ACL where the CM5 and other state of the art computers are located. Another initial site will be the LDCC for similar reasons. The third site has yet to be identified but will probably be at a user's location. Each ATM switch will have 8 or more ports that will connect to various workstations. During the first phase of the project these workstations will be SGI (Onyx), Sun (Sparc w/ S-Bus), IBM RS6000, and possibly HP (735). The switches will be interconnected with single mode fiber optics which wereinstalled as part of the LAICS project. The switches also have Ethernet and FDDI ports which allow them to operate as gateways between these networks and the Sunrise ATM Network.

Initially, the ATM switches will only support variable bit rate (VBR) service and will be operating at 100 megabits/sec. The constant bit rate (CBR) service for support of voice and video will come later as will higher data rates of 155 megabits under the Synchronous Optical NETwork specifications (SONET). The SONET aspect is an important issue because of two reasons. The first is that the Laboratory is also funding an effort to connect all Lab sites to a network called LANLnet. While the initial goal of this project is connectivity for email and file transfer purposes for all Laboratory employees, there is a vision of expanding LANLnet to take advantage of some of the technology developments from Sunrise. The

LANLnet backbone will most likely be based on the SONET standard and grow to include ATM services much like Sunrise. Therefore, it is important to focus the Sunrise network to use the technology we expect to be using in LANLnet.

The second reason is that the telephone companies from the local carriers (US West) to the long haul carriers (AT&T, MCI, Sprint, etc) are now building a SONET infrastructure to support wide area networks using ATM. Their goal is to have the ATM switch located at the central office (CO) of your telephone company rather in your computer room. The important point, however, is that the network technology will be the same for local area networks (LANs) as well as wide area networks (WANs), which is definitely not the case today. Using similar technologies in the WAN and LAN make interconnectivity "seamless" and much more robust. The much hyped "National Data Super Highway" looming on the horizon will probably use this type of technology.

Another part of the project plan is to connect to gigabit networks. Los Alamos has been developing gigabit networks based on the High Performance Parallel Interface or HIPPI for the past few years. The ACL has an access point to the HIPPI Testbed that interconnects various supercomputers having HIPPI ports together with high speed disk arrays and frame buffers. The ACL is also the entry point for the Los Alamos link to the CASA Gigabit Testbed which connects the Laboratory with supercomputer centers in San Diego (SDSC) and Pasadena (Caltech and JPL). The link between these sites is based on HIPPI but uses the SONET infrastructure of the local and long haul carriers. One goal of the networking aspect of Sunrise is to use a HIPPI-ATM gateway or router to facilitate this function. There are several issues involved in building such a device. Several vendors have proposed products to produce the router function and Los Alamos, as well as other research institutions, areinvestigating HIPPI-ATM gateways.

Current Status

The first of the 3 ATM switches has been delivered to Los Alamos along with interfaces for Sun (S-Bus), IBM-RS6000 (micro-channel), and SGI (VME bus). These first systems will operate at only 100 megabits/sec. The equipment is being assembled into an ATM testbed for initial checkout. The other two switches are being procured on an competitive procurement based on our evaluation of this first switch. A higher speed (155 megabits/sec) module for the switch is coming in a few months that will support SONET transmissions. The SONET technology is important for Sunrise connectivity to different vendor's switches and to wide area networks connections into the NII.

Remaining tasks for FY94

After setting up the ATM testbed around this first switch, an extensive evaluation and learning phase will begin since ATM networking is a new technology at the Laboratory. This evaluation will be ongoing throughout the project as different components of ATM networks become available. We especially want to evaluate new ATM switches that support the higher bandwidth SONET technology.

FY95 Tasks

The major emphasis for FY95 will be to move the ATM Network out of the "testbed" phase and into a production phase. This will involve locating the switches at the "strategic" sites, installing ATM interfaces on specific workstations and running Sunrise applications such as Transims and Telemedicine over the ATM network. Additional network support of these applications centers around ATM's ability to carry mulitmedia transmissions. This feature, called Constant Bit Rate (CBR) service, is not supported in the first offerings of ATM interfaces for workstations. A primary FY95 task will be to pursue the implementation of CBR service on the Sunrise ATM network by obtaining new interfaces from vendors or using the electronic design capabilities in CIC-5 to prototype a CBR ATM interface.

Another effort in FY95 will be to introduce a ATM/HIPPI gateway connecting the Sunrise Network to the LANL HIPPI Gigabit Network. This will be done initially with an IBM RS6000 workstation containing HIPPI and ATM interfaces.

Milestones:

April 94

Delivery of first ATM switch with interfaces for SGI, Sun, and IBM

operating at 100 Mbit/sec

May 1994

Demonstration of network capabilities with ATM switch.

July 1994

Delivery of two more ATM switches with more interfaces.

Delivery of ATM interfaces for specific Sunrise Hosts operating at

155 Mbit/sec (OC3 SONET)

September 1994

Place switches in "strategic" sites for Sunrise

Connect hosts at each site into ATM network

October 1994

Demonstrate ATM network with Sunrise applications.

January 1995

Upgrade switches and interfaces to support multimedia applications

March 1995

Demonstrate HIPPI-ATM function in Sunrise

Connect Sunrise ATM network to HIPPI gigabit network

July 1995

Demonstrate multimedia applications on Sunrise Network

Related Laboratory capabilities and investigator's expertise

There are a number of information technology projects that have been started at the Laboratory during the past year. These include the Information Architecture project, the Info2000 project, the Library without Walls project, the DoE Information Infrastructure Initiative. Each has a different emphasis and in each case we have participants of the Sunrise project who are involved in or familiar with each of these. Although the projects do not overlap, we are able to share and exchange information on a regular basis. Another project started this past year is DoE's Gas and Oil National Information Infrastructure project. We have found that many of the technologies being developed in the Sunrise project are immediately applicable to this project. The variety of investigators have considerable expertise in Information Infrastructure technology and its use in technical applications. The data-mining capabilities provided, by C-3, for example, are unmatched anywhere in the country. Dick Phillips, a world-renowned multimedia document expert, is providing the executive user interface technology to these applications. In addition, the high performance networking expertise here at Los Alamos is a world leader in the field. This effort is a natural integration of that expertise with the demands of both high performance computing and the requirements of ubiquitous service.